

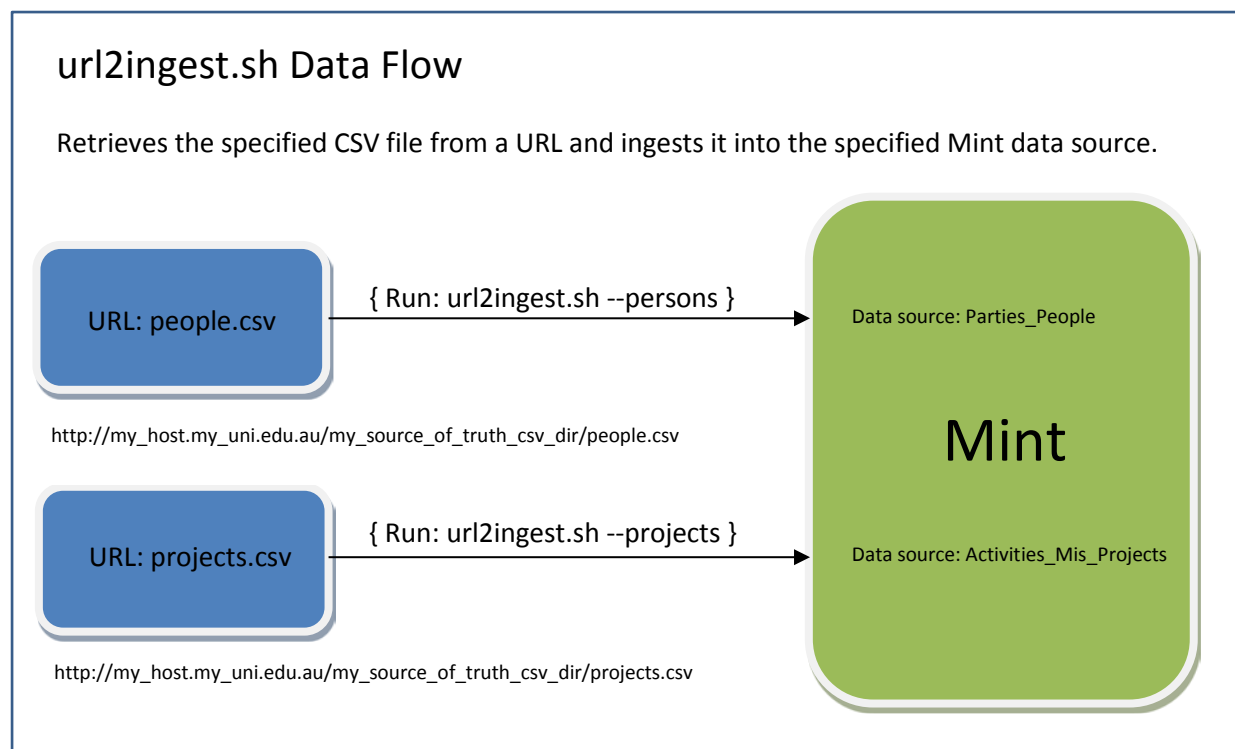
FlindersRedbox-url2ingest: Development documentation

See the INSTALL document for information regarding Application Environment and Installation.

Although the script is written in bash, it is expected that it will operate with little or no modification under sh and ksh shells.

Data Flow

Unless incremental-load is specified on the command line, there is no data transformation performed by the script. If incremental-load is specified (not shown in the diagram below) the data loaded into Mint is filtered to leave only new or updated records. The script data flow diagram is shown below.



High level algorithm

The following high level algorithm is used in the url2ingest.sh script.

```
Read command line arguments and initialise URL, data source, etc.
Get CSV file from URL and store in download directory.

IF full load argument specified THEN
    Point to full-CSV file via symlink. [1]
    Load CSV file into Mint data source.

ELSE # Incremental load argument specified
    Find the most recent full-CSV file.
    Extract new and updated CSV records (compared with most recent full-CSV file).

    IF there are any new or updated records THEN
        Point to incremental-CSV file via symlink. [1]
        Load CSV file into Mint data source.
        Backup this incremental-CSV file to history directory.
    ENDIF

ENDIF

IF the CSV-load was successful THEN
    Backup this full-CSV file to history directory.
ENDIF

IF the CSV-load was attempted (successful or not) THEN
    Backup Mint harvest.out file into log directory.
ENDIF
Clean up download directory, working directory, etc.
```

Notes

[1] A symlink from the CSV file in the Mint tree must be configured in advance to point to the appropriate TARGET_SOURCE_PATH for the data source. In addition during the running of the script, TARGET_SOURCE_PATH shall be configured to point to either the full-CSV file or the incremental-CSV file.